



## Informationswissenschaft für Linguisten

### Die Sprache des Information retrieval

Engerer, Volkmar Paul

#### *Published in:*

Deutsch als Bindeglied zwischen Inlands- und Auslandsgermanistik

#### *Publication date:*

2017

#### *Document version*

Også kaldet Forlagets PDF

#### *Citation for published version (APA):*

Engerer, V. P. (2017). Informationswissenschaft für Linguisten: Die Sprache des Information retrieval. In S. Nefedov, L. Grigorieva, & B. Bock (Eds.), *Deutsch als Bindeglied zwischen Inlands- und Auslandsgermanistik: Beiträge zu den 23. GeSuS-Linguistik-Tagen in Sankt Petersburg, 22.-24. Juni 2015* (pp. 93-101). Verlag Dr. Kova. Sprache und Sprachen in Forschung und Anwendung (SiFA) Vol. 5

**Internetportale, gesehen am 23.01.2015**

<http://www.radio.cz/de/rubrik/kaleidoskop/streit-um-ferdinand-porsche-spaltet-tschechischen-ort-vratislavice>  
<http://lehrerfortbildung-bw.de/bs/bsa/bgym/lehrga>

# **Informationswissenschaft für Linguisten. Die Sprache des Information retrieval**

Volkmar Engerer

## **1 Einleitung**

In diesem Beitrag<sup>1</sup> möchte ich das sprachliche und kommunikative Potenzial der Informationswissenschaft (im Folgenden abgekürzt mit „IW“) ausloten und dem linguistisch gebildeten Leser ein anderes, aber doch verwandtes Wissensgebiet vermitteln. Das Modell des Information retrieval („IR“; auf Deutsch etwas umständlich „Informationsrückgewinnung“), das ich im Folgenden näher erläutern werde, ist aus meiner Sicht die Brücke von der Informations- zur Sprachwissenschaft, da dieses Modell der Informationssuche in hohem Maße von der Sprache und ihrer kommunikativen Verwendung lebt. Nicht immer sind sich Informationswissenschaftler über die Reichweite der sprachlichen Fundierung ihres Gegenstandes klar, und ebenso wenig kann das IR als linguistisch relevantes und praktisch anwendbares Forschungsgebiet bei Sprachwissenschaftlern als bekannt vorausgesetzt werden. Dieser Beitrag zieht die sprachlichen und kommunikativen Verbindungslinien zwischen dem IR und der Linguistik und möchte dadurch einen Beitrag zu einem besseren gegenseitigen Verständnis der beiden Disziplinen leisten.

Einstieg in die Problematik ist die einfache Frage, was die IW ist bzw. macht. Diese Frage wird dann zu drei wissenschaftstheoretischen Zugängen (realistische, essentielle und problemlösungsorientierte) führen, von denen, wie ich zeige, der letztgenannte vorzuziehen ist (zumindest in Bezug auf seine Affinität zu Sprache und Kommunikation). Das mit dem problemlösungsorientierten Ansatz korrespondierende Szenario des IR wird als „Kern“ der IW motiviert und abschließend werden die beiden Äste des IR, der benutzerbezogene und der dokumentbezogene, auf ihr linguistisch relevantes Potenzial hin untersucht. Ich schließe ab mit einem kleinen Ausblick.

## **2 Was ist/ macht die Informationswissenschaft?**

In informationswissenschaftlichen Einführungen, Handbüchern und Nachschlagewerken (z.B. Stock/Stock 2013, Bawden/Robinson 2012, Bates 2010) wird generell festgehalten, dass es keine allgemein akzeptierte Definition der IW gibt.

---

<sup>1</sup> In gewissem Sinne ist dieser Beitrag eine Fortsetzung meines Artikels „Indexierungstheorie für Linguisten. Zu einigen natürlichsprachlichen Zügen in künstlichen Indexsprachen“, der in den Akten vom Arbeitstreffen der Gesus 2013 in Freiburg/Breisgau erschienen ist (Engerer 2014). Hier wie dort versuche ich, als „gelernter“ Linguist und „angelernter“ Informationswissenschaftler die beiden Disziplinen einander näherzubringen.

Sie sei ein junges Fach, ca. 50 Jahre alt, und in hohem Masse mit anderen Disziplinen verbunden, also ganz grundlegend interdisziplinär. Dementsprechend kann die Frage, was die IW ist, ganz verschieden angegangen werden. Ich möchte im Folgenden drei Arten, diese Frage zu beantworten, unterscheiden.

Fragen wie „Was sind die Aufgaben der IW?“, „Wie hat sich die IW historisch entwickelt?“ oder „Mit welchen anderen Disziplinen unterhält die IW Beziehungen?“ sind **realistisch** in dem Sinne, dass sie ihren Ausgangspunkt im Status Quo und der gegebenen historischen Entwicklung der IW nehmen. Der realistische Standpunkt impliziert eine quasi aufzählende, extensionale Vorgehensweise bei der Charakterisierung einer Disziplin und ist in Bezug auf die Konstituierung der IW als Disziplin in dem Sinne nicht-unitarisch, da von realistischen Statements nicht unmittelbar zukunftsbezogene und vereinigende Kriterien für die IW als kohärente Disziplin abgeleitet werden können. Ein anderer Zugang spiegelt sich in der Hinterfragung von wesentlichen informationswissenschaftlichen Konzepten wie ‚Information‘, ‚Wissen‘ und ‚Dokument‘ wider. Fragen dieses Typs zielen auf einen semantischen „Kern“, einen intensionalen Inhalt des Ausdrucks *Informationswissenschaft*. Dieser Zugang kann als „**essentiell**“ gelten. Die essentielle Fundierung einer Disziplin fragt nach dem konzeptuell Verbindenden und umgeht damit zufällige und kontingente Referenten, die ein notwendiges und typisches Nebenprodukt einer jeden realistischen Analyse sind. Die essentielle Strategie ist auf die Stabilisierung als eigenständige Wissenschaft gerichtet und damit unitarisch für die Formierung der IW als Wissenschaft. Eine dritte Möglichkeit der Annäherung an die IW ist **problemorientiert** und hat den Informationssuchprozess zum Gegenstand. Aus dieser Perspektive untersucht die IW die Repräsentation, die Speicherung und die Versorgung sowie die Suche nach und die Wiederfindung von relevanten Dokumenten und relevantem Wissen (vgl. „Informationsrückgewinnung“) (Stock/Stock 2013: 3). Hiernach ist das definierende Szenario der IW grundsätzlich problematisch, da es auf einer Defizithypothese auf Seiten des Benutzers eines Wissenssystems beruht. Auch der problemorientierte, prozessuale Zugang zur Disziplin ist, wie der essentielle, von unitarischem Charakter, doch weniger auf die semantische Integrität der Referenten gerichtet, als vielmehr auf einen uniken prozessualen referentiellen Gegenstand und dessen pragmatische Zusammenhänge, z.B. Akteure (Benutzer) und deren psychologische Zustände (Wissensdefizit), Patiens-Objekte (Dokumente), die von Akteuren als Informationsträger angesehen werden, usw.



## 2.1 Essentielle Zugänge

Essentielle Zugänge zur IW spiegeln sich wider in der Frage „Welche Rolle spielen Information, Wissen und Dokument in der IW?“ und zielen auf die sprachliche Bedeutung der morphologischen Komponenten des Wortes *Informationswissenschaft*, also *information* und *wissenschaft*, ab. Andere essentielle Schlüsselbegriffe schließen sich in deduktiver Weise an die beiden dominierenden Konstituenten an, so ergibt sich z.B. ‚Wissen‘ als ‚statischer Gegenstand dynamischer Information‘ oder ‚Dokument‘ als ‚materialisierte Form von Information‘ (s. dazu Stock/Stock 2013: 22–39 und die dort zitierte Literatur). Ziel des essentiellen Ansatzes ist damit die „Essenz“ und der inhaltliche Kern einer Wissenschaft.

Es ist in der theoretischen IW charakteristisch, dass Information in Zusammenhang mit Wissen betrachtet wird, oft konzeptualisiert als ‚Information ist Wissen in Bewegung‘ (vgl. Kuhlen 1995). Die Verbindung von Wissen und Information wird im *Handbook of Information Science* (Stock/Stock 2013) an Poppers Theorie der drei Welten gezeigt (Popper 1994):

- Welt 1: physikalische Welt
- Welt 2: Welt der bewussten Erfahrung (mental, subjektives Wissen)
- Welt 3: Welt des „Inhalts“ (objektives, dokumentiertes Wissen in z.B. Artikeln und Büchern, existiert unabhängig von Menschen).

Die Welten 2 und 3 hängen eng zusammen, indem subjektives Wissen auf dokumentiertem Wissen der Welt 3 beruht, und umgekehrt, objektives Wissen sich aus Welt-2-Wissen heraus entwickelt. Beim Übergang von Welt 3 zu Welt 2 und bei Wissensprozessen innerhalb Welt 2 kommt dann Information ins Spiel. Information ist die **Form**, die in Welt 2 verstanden wird. Wissen als solches kann nicht transportiert werden, Wissen ist angewiesen auf Information als Träger (Stock/Stock 2013: 24). Daher auch die oben angesprochene Formel „Information ist Wissen in Bewegung“.

Einen rein semantischen Blickwinkel auf „Information“ legt z.B. Buckland an (Buckland 1991, s. auch Stock/Stock 2013: 25). Buckland schließt von den drei „Bedeutungsaspekten“ des Ausdrucks *Information*, in einem praktisch-ontologischen Gegenzug sozusagen, auf die realen Objekte der IW:

- Information als Prozess (Gebrauch des Verbs *informieren* wie in der Äußerung „Peter informiert Helga“)
- Information als Wissen (Information transportiert Wissen, so dass Helga aus dem obigen Beispiel nun die besagte Information „hat“)
- Information als Ding (die Fixierung von Information in „informativen Dingen“, z.B. Dokumenten; hier hat Helga, wieder in Fortführung obigen Beispiels, einen entsprechenden Brief von Peter bekommen).

Buckland baut also ontologische Statements, die wiederum Input für eine essentielle Grundlegung der IW sind, auf einer semantischen Analyse des Ausdrucks *Information* auf.

Eine lange Diskussion ist in der IW übrigens über ihren wissenschaftlichen Status geführt worden (vgl. die Komponente *-wissenschaft* in *Informationswissenschaft*). „Wissenschaft“ und „wissenschaftlich“ werden und wurden oft im Sinne einer naturwissenschaftlichen und positivistischen Methodologie verstanden, was es, so die Kritiker, für die Mainstream-IW schwer machte, bedeutungsbezogene Phänomene zu verstehen (vgl. Tredinnick 2006).

## 2.2 Realistische Zugänge

Ein realistischer Zugang zur IW stellt sich Fragen wie „Was sind die Aufgaben der IW?“, „Wie hat sich die IW historisch entwickelt?“ oder „Mit welchen anderen Disziplinen unterhält die IW Beziehungen?“. Hierzu ist zu sagen, dass sich die IW in enger Zusammenarbeit mit der Computerwissenschaft und der Informationsverarbeitung entwickelt hat. Des Weiteren besteht ein traditionell enges Verhältnis zur Bibliothekswissenschaft (deshalb englisch LIS = Library and Information Science), aber auch zur Lerntheorie und Pädagogik (u.a. in virtuellen Lehrplattformen). Historisch hat sich die IW in den 50er Jahren entwickelt, also als die Computer- und Datenrevolution fühlbar wurde, und sie ist immer noch eine stark angewandte Disziplin.

## 2.3 Problemorientierte Zugänge

Laut Stock/Stock (2013: 3) untersucht die IW die Repräsentation, die Speicherung und die Versorgung sowie die Suche nach und die Wiederfindung von relevanten Dokumenten und relevantem Wissen. Bestandteile dieser Definition sind (Stock/Stock 2013: 3–5):

- **Repräsentation:** Alles Wissen in Dokumenten (z.B. wissenschaftliche Artikel) ist kondensiert in verkürzten Texten (Abstracts) oder Deskriptoren/Schlagworten, die der Informationsfiltrierung dienen
- **Speicherung und Versorgung:** Strukturierung einer Datenbank (eines Bibliothekskataloges, ...) in der Weise, dass die Dokumente einfach wiedergefunden werden können und inhaltlich zugänglich sind
- **Suche:** Bringt den Benutzer im Prozess der Deckung seines Informationsbedarfs ins Spiel; untersucht werden z.B. die verwendeten Suchausdrücke in Bezug auf die jeweiligen Suchwerkzeuge und die Art und Weise, wie Benutzer die gefundene Information weiterverarbeiten
- **Informationsrückgewinnung/IR:** Entwicklung von Systemen für die Erforschung von Wissen wie Internet-Suchmaschinen oder Bibliothekskataloge



- **Relevanz:** Grundbegriff der IS; nicht jegliche Information ist Gegenstand von Suchen, sondern ausschließlich solche Informationen, welche dem Suchenden dabei helfen, seinem Informationsbedarf abzuhelpfen
- **Dokumente:** Sowohl textliche als auch nicht-textliche Objekte wie Bilder, Musik, Videos, ...
- **Wissen:** Statischer Inhalt, fixiert in Dokumenten oder gespeichert in einem mentalen Gedächtnis. Wissen kann digital (WWW), materiell (Bücher auf einem Bibliotheksregal) oder psychisch-mental sein (Gehirn eines Angestellten einer Firma). Information ist dagegen dynamisch.

Das definierende Szenario der IW ist aus dieser Perspektive grundsätzlich problematisch und beruht auf einer Defizithypothese auf Seiten des Benutzers eines Wissenssystems, wie es in Belkins berühmten ASKs, den „anomalous states of knowledge“, auf den Punkt gebracht wird (Belkin 1980).

#### 2.4 Abwägung essentieller, realistischer und problemorientierter Zugänge

In Bezug auf den essentiellen Zugang scheint es schwierig zu sein, semantisch basierte Erkenntnisse über *Information*, *Wissen* und *Dokument* mit praktischen Problemstellungen der IW kombinieren zu können. Die realistische Methode dagegen hat wohl dieses praktische Potenzial, ist aber rein deskriptiv ausgerichtet und verfügt über kein Potenzial für die IW als Disziplin in einem unitarischen Sinne. Problemorientierte Strategien scheinen mir dagegen der vorzuziehende Umriss einer sowohl praktischen als auch theoretisch fundierten IW zu sein. Hier einige der Eigenschaften des problemorientierten Blickwinkels, die in diesem Sinne positiv zu werten sind:

- Das Szenario ist qua problematischer Grundkonstellation lösungsfördernd; dies nährt sowohl den intellektuellen als auch den praktischen Impetus zur wissenschaftlichen Handlung
- Das Szenario ist allgemein und trotzdem spezifisch, und kann auf viele ähnliche Situationen angewendet werden; es wird der Domänenvielfalt in der IW gerecht und findet Anwendung z.B. in professionellen Domänen (z.B. der Entscheidungsfindung), den sozialen Medien, Alltagsdomänen wie der Freizeit oder Informationsbeschaffung des Bürgers im öffentlichen Sektor, in Wissenschaft und Forschung, Ausbildung, Militär, ...
- Das Szenario ist benutzerintegrierend: Der Benutzer, mit allen seinen individuellen, psychologischen und kulturellen Kennzeichen geht durch das Konzept des Informationsbedarfes und der Spezifikation relevanter Information als zentraler Bestandteil in das Szenario ein; der rein technologischen Komponente wird dadurch ein humanistischer Mitspieler an die Seite gestellt
- Das Szenario fordert interdisziplinäre Zusammenarbeit: Wissen über die einzelnen Komponenten ist auf eine Vielzahl von Wissenschaften verteilt, z.B. die Computervissenschaft, Psychologie, Soziologie, Design, Philosophie, Kulturwissenschaften und mehr
- Das Szenario fordert durch seine handlungsmäßig verbundenen Komponenten zusammenhängende Problemstellungen heraus; anders gesagt, es schafft einen integrierenden Rahmen, welcher der Erforschung von Einzelproblemen Kohärenz und Zusammenhang verleiht.

Die problemorientierte Komponente in der IW ist im sog. interaktiven „Information retrieval“-Modell implementiert (Baeza-Yates/Ribeiro-Neto 2011, Ruthven/Kelly 2011, Warner 2010, Borlund 2013), dessen Elemente unten zu sehen sind.

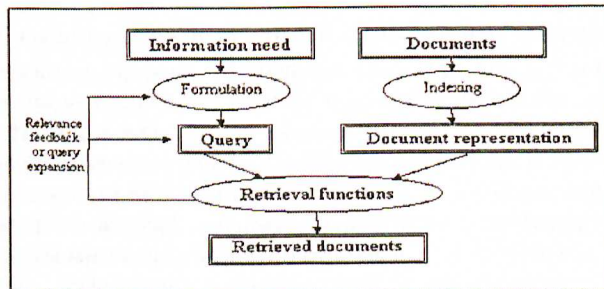


Abb. 1. Modell des interaktiven IR (nach Gharaibeh/Gharaibeh 2012)





semantisch sehr viel reicher strukturiert als die Lexika des IR (genannt „Thesauri“). In noch höherem Maße kann dies für die Syntax des Benutzers im Vergleich zu maschinengerechten Suchausdrücken gelten. Die kommunikative Situation kann weiterhin dadurch verkompliziert sein, dass ein „intermediate“ zwischen Benutzer und Suchmaschine geschaltet ist, also ein Informationsspezialist (z.B. Bibliothekar), der zwischen dem Benutzer und dem System vermittelt. Hier sind komplexe Verstehensprozesse beteiligt, die z.B. das interaktive Hinarbeiten der Partner zum „Gemeintem“ des Benutzers, seinem Informationsbedarf, und die „Übersetzung“ des „Verstandenen“ in einen semantisch äquivalenten Suchausdruck durch den Informationsspezialisten mit sich ziehen. Eine weitere, etwas anders geartete Kommunikationssituation ist in der Schnittstelle zwischen Benutzer und Maschine gegeben, welche auf der Maschinenseite mindestens zwei Repräsentationsebenen involviert, eine maschinennahe Repräsentation der Zustände (z.B. „Kein Dokument gefunden“) und ihre Repräsentation im Interface des Computers, welche auf die zukünftigen sprachlichen Züge des Benutzers („Sucheingaben“) Einfluss nehmen will (z.B. die Aufforderung an den Benutzer, einen breiteren Suchbegriff zu verwenden). Diese Ebene der Kommunikation ist in obiger Abbildung durch „relevance feedback“ und „query extension“ repräsentiert.

Sprachliche Repräsentationsebenen, die auf dem rechten Ast angesiedelt sind, betreffen ganz zentral den Prozess der textuell-sprachlichen Repräsentation von Dokumentinhalten. Die Eigenschaft von Texten, über etwas zu handeln, ein Thema (oder mehrere Themen) zu besitzen, auf die dann in einer thematischen Suchabfrage zugegriffen werden kann, heißt in der IW „aboutness“ (Beghtol 1986, Bawden/Robinson 2012: 113, Frohmann 1990: 83, Blair 1992). Das Erfassen und die Beschreibung des Themas/der Themen eines Textes (sowohl durch Mensch als auch Maschine) wird in der IW als „Indexierung“ bezeichnet (Svenonius 2000, Lancaster 2003, Fugmann 2002, Mai 1999). Indexierung betrifft aus linguistischer Sicht die Analyse und Interpretation der Textbedeutung und die Repräsentation dieser Textbedeutung mit Hilfe eines kontrollierten, maschinennahen Vokabulars. Des Weiteren geschieht diese Repräsentation der Textbedeutung mit Hinblick auf den Sprachgebrauch und das Idiom eines zukünftigen Benutzers („benutzergerichtete Indexierung“). Nicht zuletzt sollte die syntaktische Kombination der Indexterme in einem Beschreibungsausdruck (normalerweise eine Kette von sog. „Deskriptoren“) sowohl der ontologischen Komposition des Textreferenten „in der Welt“ entsprechen als auch auf mögliche Übereinstimmungen mit der Benutzersyntax in Abfragen gerichtet sein.

Das IR-Modell eröffnet eine Reihe interessanter Problemstellungen sprachlicher und kommunikativer Art, wie schon diese wenigen Bemerkungen zeigen.

#### 4 Ausblick

Ich meine, dass die Möglichkeiten einer Zusammenarbeit zwischen Informationswissenschaftlern und Linguisten noch lange nicht ausgenutzt sind. Computerlinguisten und andere angewandt orientierte Linguisten haben schon immer ihre Chance in der IW gesucht, allgemeine, mehr traditionell und theoretisch arbeitende Sprachwissenschaftler dagegen scheinen hier größere Berührungsängste zu haben. Ich hoffe, mit diesem Beitrag gezeigt zu haben, dass dazu kein Grund besteht. Linguisten haben der Informationswissenschaft viel zu sagen, und viele Elemente in dem besprochenen IR-Modell überdecken nur sprachliche und kommunikative blackboxes, in denen sich komplexe und linguistisch interessante Prozesse verstecken. Licht in diese blackboxes zu bringen würde nicht nur die IW weiterbringen, auch Linguisten könnten von dem handfesten, problemorientierten Zugang des IR-Szenarios profitieren. Viel von dem, was in der Mainstream-Linguistik erst langwierig problematisch geredet und theoretisiert werden muss, ist in der IW vorausgesetzt und muss nicht erst konstruiert werden.

#### Literatur

- Baeza-Yates, Ricardo/Ribeiro-Neto, Berthier 2011: *Modern information retrieval: the concepts and technology behind search*. 2nd ed. New York: Addison Wesley.
- Bates, Marcia J. 2010: *Encyclopedia of library and information sciences*. 3rd ed. New York: Marcel Dekker.
- Bawden, David/Robinson, Lyn 2012: *An introduction to information science*. London: Facet.
- Beghtol, Clare 1986: "Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents". *Journal of Documentation* 42(2), 84–113.
- Belkin, Nicholas J. 1980: "Anomalous states of knowledge as a basis for information retrieval". *Canadian Journal of Information. Science-Revue Canadienne Des Sciences De L'Information* 5 (May), 133–143.
- Blair, David C. 1992: "Information Retrieval and the Philosophy of Language". *The Computer Journal* 35 (3), 200–207.
- Borlund, Pia 2013: "Interactive Information Retrieval: An Introduction". *Journal of Information Science Theory and Practice* 1(3), 12–32.
- Buckland, Michael K. 1991: "Information as thing". *Journal of the American Society for Information Science* 42 (5), 351–360.
- Frohmann, Bernd 1990: "Rules of indexing: a critique of mentalism in information retrieval theory". *Journal of Documentation* 46 (2), 81–101.
- Fugmann, Robert 2002: "The complementarity of natural and index language in the field of information supply: an overview of their specific capabilities and limitations". *Knowledge Organization* 29 (3/4), 217–230.
- Gharaibeh, Islah K./Gharaibeh, Natheer K. 2012: "Towards Arabic Noun Phrase Extractor (ANPE) Using Information Retrieval Techniques". *Software Engineering* 2 (2), 36–42.
- Kuhlen, Rainer 1995: *Informationsmarkt: Chancen und Risiken der Kommerzialisierung von Wissen*. Konstanz: Universitätsverlag Konstanz.

- Lancaster, F. Wilfrid 2003: *Indexing and abstracting in theory and practice*. 3rd. ed. London: Facet.
- Mai, Jens-Erik 1999: "Deconstructing the Indexing Process". *Advances in Librarianship* 23, 269–298.
- Popper, Karl R. 1994: *Objective knowledge, an evolutionary approach*. Rev. ed., 7. impr. Oxford: Oxford University Press.
- Ruthven, Ian/Kelly, Diane 2011: *Interactive information seeking, behaviour and retrieval*. London: Facet.
- Stock, Wolfgang G./Stock, Mechtild 2013: *Handbook of information science*. Berlin: de Gruyter Saur.
- Svenonius, Elaine 2000: *The intellectual foundation of information organization*. Cambridge, Mass: MIT Press.
- Tredinnick, Luke 2006: *Digital information contexts: theoretical approaches to understanding digital information*. Oxford: Chandos Publishing.
- Warner, Julian 2010: *Human information retrieval*. Cambridge, Mass.: MIT Press.